Computationally Harnessing Wikipedia’s Knowledge

Shilad Sen
http://www.shilad.com
ssen@macalester.edu
1. WikiBrain

2. WikiBrain-driven projects
Shilad’s Macademia profile - http://macademia.macalester.edu
Metaxas, P. Takis

affiliations: Wellesley College
dep: Computer Science
email: pmetaxas@wellesley.edu
interests: Web Science, trust, parallel computing, privacy, computer science education, misinformation, computational social science, elections, information retrieval, predictive analytics, social networks, computer science, politics, Web Spam, multimedia, network analysis, medical software, web search, propaganda

related to Shilad Sen by:
- politics
- Web Spam (similar to web2.0)
- web search (similar to web2.0)
- network analysis (similar to social networking)
- social networks (similar to social networking)
- predictive analytics (similar to statistics, data mining)
- computational social science (similar to collaborative computing)
WP:Clubhouse? An Exploration of Wikipedia’s Gender Imbalance

Shyong (Tony) K. Lam\textsuperscript{1}, Anuradha Uduwage\textsuperscript{1}, Zhenhua Dong\textsuperscript{2}, Shilad Sen\textsuperscript{3}, David R. Musicant\textsuperscript{4}, Loren Terveen\textsuperscript{1}, John Riedl\textsuperscript{1}

\textsuperscript{1}GroupLens Research  
Dept. of Computer Science and Engineering  
University of Minnesota  
Minneapolis, Minnesota  
\{lam, uduwage, terveen, reidl\}@cs.umn.edu

\textsuperscript{2}Dept. of Information Technical Science  
Nankai University  
Tianjin, China  
dongzhh@mail.nankai.edu.cn

\textsuperscript{3}Math, Statistics, and Computer Science Dept.  
Macalester College  
St. Paul, Minnesota  
ssen@macalester.edu

\textsuperscript{4}Dept. of Computer Science  
Carleton College  
Northfield, Minnesota  
dmusican@carleton.edu

\textbf{ABSTRACT}

Wikipedia has rapidly become an invaluable destination for millions of information-seeking users. However, media reports suggest an important challenge: only a small fraction of Wikipedia’s legion of volunteer editors are female. In the current work, we present a scientific exploration of the gender imbalance in the English Wikipedia’s population of editors. We look at the nature of the imbalance itself, its effects on the quality of the encyclopedia, and several conflict-related factors that may be contributing to the gender gap. Our findings confirm the presence of a large gender gap among editors and a corresponding gender-oriented disparity in the content of Wikipedia’s articles. Further, we find evidence hinting at a culture that may be resistant to female participation.

\textbf{Categories and Subject Descriptors}

H.3.4 [Information Systems]: Systems and Software—Information systems software—User interfaces
Computational sociology

From Wikipedia, the free encyclopedia

Computational sociology is a branch of sociology that uses computer simulations, artificial intelligence, network analysis, computational sociology develops and tests social interactions.[1] It involves the understanding of social agents, the interaction of agents, and the aggregate.[2] Although the subject matter and methodologies may be similar to social network analysis, web search, propaganda, fraud, and politics, computational sociology is an emerging field in social science.

Related to Shilad Sen by:
- politics
- Web Spam (similar to web2.0)
- web search (similar to web2.0)
- network analysis (similar to social networking)
- social networks (similar to social networking)
- predictive analytics (similar to statistics, data mining)
- computational social science (similar to collaborative computing)

Collaborative Computing Project for NMR

From Wikipedia, the free encyclopedia

The Collaborative Computing Project for NMR (CCPN) is a project that brings together an international community involved in NMR spectroscopy, especially those who work with existing NMR software via a common data standard and provide a forum for the exchange of ideas and the development of new techniques. CCPN was initially started in 1999.

Contents [hide]
1 The Collaborative Project for the NMR Community
2 NMR Data Standards
Brent Hecht

Photo by Brent Hecht, CC-BY-SA 3.0
conspiracy theory
The challenge of Wikipedia algs

Wikipedia is crucial to NLP, AI, and geospatial algs.

but...

Wikipedia is big.

Wikipedia is messy.

Robust implementations of algorithms are rare.

Research is difficult to reproduce.
Enter WikiBrain

Mission: Democratize access to state-of-the-art Wikipedia algorithms and technologies.

Audience: Programmers with basic Java (for now).

Focus: Core data structures, AI, NLP, Geospatial.

Design goals: Fast, flexible, easy to use (3rd gen).
Resources related to Shilad's 2014 OpenSym talk:

- Talk slides
- WikiSym 2014 paper
- Source files: Quickstart.java, TranslateConcept.java, SimilarMovies.java, CountryPageViews.java, SimpleToblersEvaluator.java, CategoryViews.java

WikiBrain's busy thinking up its first public release. Please be patient while we fine tune our APIs and complete our documentation. Ask us questions at the WikiBrain google group!

The WikiBrain Java library enables researchers and developers to incorporate state-of-the-art Wikipedia-based algorithms and technologies in a few lines of code.

**WikiBrain is easy to use.** Wikipedia data can be downloaded, parsed, and imported into a database by running a single command. WikiBrain allows you to incorporate state-of-the-art algorithms in your Java projects in just a few lines of code.

**WikiBrain is multi-lingual.** WikiBrain supports all 267 Wikipedia language editions, and builds a concept-map that connects an article in one language to the same article in another language.

**WikiBrain is fast.** WikiBrain uses single-machine **parallelization** (i.e. multi-threading) for all computationally intensive features. While it imports data into standard SQL databases (h2 or Postgres), it builds optimized local caches for critical data.

WikiBrain integrates a variety of specific algorithms and datasets in one framework, including:

- **Semantic-relatedness** algorithms that measure the strength of association between two concepts such as "racecar" and "engine."
- **GeoSpatial** algorithms for spatial Wikipedia pages like Minnesota and the Eiffel Tower.
- **Wikidata**: Support for structured Wikidata "facts" about articles.
- **Pageviews**: Public data about how often Wikipedia pages are viewed with hourly precision.

```
// An example program
```
** ALL DIAGNOSTIC TESTS SUCCEEDED! **

Rough estimate of download size: 1236.0 MBs
   This may be an over-estimate if some files have already been downloaded.
   Time on dial-up (50kbs): 4120.0 minutes
   Time on Broadband (1Mbs): 206.0 minutes
   Time on Broadband (10Mbs): 20.6 minutes
   Time on Broadband (100Mbs): 2.1 minutes
   stage download will download about 576.0 about MBs
   stage concepts will download about 660.0 about MBs

Completion time estimate: 7.1 minutes (NOT including download time)
   stage fetchlinks: 0.0 minutes
   stage download: 0.0 minutes
   stage concepts: 6.7 minutes
   stage sr: 0.3 minutes

Disk space is okay. (need 0.838 G Bs, have 39.276 G Bs)
   Warning: Available disk space may be INACCURATE if you have multiple drives.
   stage fetchlinks: 1.2 MBs
   stage download: 576.0 MBs
   stage concepts: 41.1 MBs
   stage sr: 240.0 MBs

Amount of memory allocated for the JVM is okay
   memory required: 3.0 GB
   memory allocated: 3.8 GB

Connection to database succeeded. Active configuration:
   username: "sa"
   url: "jdbc:h2://$(baseDir)/db/h2;LOG=0;CACHE_SIZE=65536;LOCK_MODE=0;UNDO_LOG=0;MAX_OPERATION_MEMORY=10000000"
   partitions: "default"
   connectionsPerPartition: 2
   driver: "org.h2.Driver"
   password: ""

Beginning import process in 20 seconds...
** ALL DIAGNOSTIC TESTS SUCCEEDED! **

Rough estimate of download size: 1236.0 MBs
This may be an over-estimate if some files have already been downloaded.
Time on dial-up (50kbs): 4120.0 minutes
Time on Broadband (1Mbs): 206.0 minutes
Time on Broadband (10Mbs): 20.6 minutes
Time on Broadband (100Mbs): 2.1 minutes
stage download will download about 576.0 about MBs
stage concepts will download about 660.0 about MBs

Completion time estimate: 7.1 minutes (NOT including download time)
stage fetchlinks: 0.0 minutes
stage download: 0.0 minutes
stage concepts: 6.7 minutes
stage sr: 0.3 minutes

Disk space is okay. (need 0.838 GBs, have 39.276 GBs)
Warning: Available disk space may be INACCURATE if you have multiple drives.
stage fetchlinks: 1.2 MBs
stage download: 576.0 MBs
stage concepts: 41.1 MBs
stage sr: 240.0 MBs

Amount of memory allocated for the JVM is okay
memory required: 3.0GB
memory allocated: 3.8GB

Connection to database succeeded. Active configuration:
username: "sa"
url: "jdbc:h2://.../db/h2;LOG=0;CACHE_SIZE=65536;LOCK_MODE=0;UNDO_LOG=0;MAX_OPERATION_MEMORY=
100000000"
partitions: "default"
connectionsPerPartition: 2
driver: "org.h2.Driver"
password: ""

Beginning import process in 20 seconds...
## Import times for core data:

<table>
<thead>
<tr>
<th>language</th>
<th># articles</th>
<th># links</th>
<th>runtime*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple English</td>
<td>102K</td>
<td>6M</td>
<td>8 min</td>
</tr>
<tr>
<td>German</td>
<td>1.9M</td>
<td>96M</td>
<td>210 min</td>
</tr>
<tr>
<td>English</td>
<td>4.6M</td>
<td>470M</td>
<td>640 min</td>
</tr>
<tr>
<td>25 largest</td>
<td>25M</td>
<td>1,670M</td>
<td>3163 min</td>
</tr>
</tbody>
</table>

*Additional time required for SR, geospatial, wikidata.
Quickstart.java

```java
public static void main(String args[]) throws Exception {
    // Prepare the environment
    Env env = EnvBuilder.envFromArgs(args);

    resolution of Apple
    Apple Inc. (simple): 0.5
    Apple (simple): 0.2769231
    Apple Records (simple): 0.2
    App Store (iOS) (simple): 0.015384615
    Apple Corps (simple): 0.0076923077

    // show the closest pages
    System.out.println("resolution of apple");
    if (resolution == null) {
        System.out.println("no resolution !");
    } else {
        for (LocalId p : resolution.keySet()) {
            Title title = pageDao.getId(p).getTitle();
            System.out.println("\t" + title + ": " + resolution.get(p));
        }
    }
}
```
Core data structures

Graphs: link, category, redirect.

Article text: wikitext and plaintext.

Full text search using Lucene.

Highly optimized disk and memory caches.
Multilingual
All languages**
Concept alignment
Universal links

Apple in other languages:
Bosnian: Jabuka
Scots: Aiple
Welsh: Afal
Icelandic: Epli
Hindi: सेब
Simple English: Apple

Pageview module

Num of page views for a requested date range.

On August 14, 2014:

Top pageviews in English
1. Main Page (en) (nviews=9961795)
2. Robin Williams (en) (nviews=312002)
3. Parkinson's disease (en) (nviews=132250)
4. Webserver directory index (en) (nviews=111069)
5. Independence Day (India) (en) (nviews=93255)
6. Java (en) (nviews=89945)
7. Lauren Bacall (en) (nviews=77279)
8. Ebola virus disease (en) (nviews=73883)
Wikidata module

Over 40M statements about 15M concepts.

What does WikiBrain know about Berlin?

+ 100 more statements

Who was born in Berlin?

+ 2000 more people in EN
Semantic relatedness module

\[ \text{similarity}(x, y) \]
\[ \text{mostSimilar}(x) \]
\[ \text{cosimilarity}(x_1, x_2, \ldots x_n) \]

\[ \text{mostSimilar}("Berlin"):\]
1. Munich
2. Hamburg
3. Vienna

\[ \text{mostSimilar}("Berlin", <all movie ids>):\]
1. The Wall (1962 film)
2. The Tunnel (2001 film)
3. The Road to the Wall
Named Entity Recognition

resolve(phrase)
resolve(phrase, context)
wikify(text)

resolve("Apple"):  
1. Apple Inc. (simple): 0.50  
2. Apple (simple): 0.28  
3. Apple Records (simple): 0.20

wikify("Wikipedia is a free-access…"):  
Wikipedia is a free-access, free-content Internet encyclopedia, supported and hosted by the non-profit Wikimedia Foundation.
Geospatial module

Built on PostGIS and OpenGeo.

Layers connected to articles:

• Wikidata (coordinate points).
• Country (polygons from NaturalEarth).
• State (polygons from NaturalEarth).

Q: How many kms separate Berlin and Alaska?
Q: How many countries separate Berlin and Shanghai?
Q: What articles are about places in Minnesota?
Spatial article views by country (EN)
public static void main(String args[]) throws ConfigurationException, DaoException {
    // Configure environment
    Env env = EnvBuilder.envFromArgs(args);
    final PageViewDao viewDao = env.getConfigurator().getPageViewDao().class;
    final LocalPageDao pageDao = env.getConfigurator().getLocalPageDao().class;
    final SpatialDataDao spatialDao = env.getConfigurator().getSpatialDataDao().class;
    final Language lang = env.getDefaultLanguage();
    final UniversalPageDao conceptDao = env.getConfigurator().getUniversalPageDao().class;
    final DateTime start = new DateTime(2014, 8, 14, 11, 0, 0);
    final DateTime end = new DateTime(2014, 8, 14, 23, 0, 0);
    viewDao.ensureLoaded(start, end, env.getLanguages());

    // Build universal id -> country shape and local page -> shape
    Map<Integer, Geometry> conceptShapes = spatialDao.getAllGeometriesInLayer("country");
    final Map<LocalPage, Geometry> countryShapes = new HashMap<LocalPage, Geometry>();
    for (int conceptId : conceptShapes.keySet()) {
        int pageId = conceptDao.getById(conceptId).getLocalId(lang);
        LocalPage page = pageDao.getById(lang, pageId);
        if (page != null) {
            countryShapes.put(page, conceptShapes.get(conceptId));
        }
    }

    // Initialize view count by country
    final Map<LocalPage, Integer> views = new ConcurrentHashMap<LocalPage, Integer>();
    for (LocalPage p : countryShapes.keySet()) views.put(p, 0);

    final Map<Integer, Geometry> conceptPoints = spatialDao.getAllGeometriesInLayer("wikidata");
    ParallelForEach.loop(conceptPoints.keySet(), new Procedure<Integer>() {
        @Override
        public void call(Integer conceptId) throws Exception {
            LocalPage country = findCountry(countryShapes, conceptPoints.get(conceptId));
            int pageId = conceptDao.getLocalId(lang, conceptId);
            if (country == null || pageId < 0) return; // probably in the ocean or outer space
            int n = viewDao.getNumViews(lang, pageId, start, end);
            views.put(country, views.get(country) + n);
        }
    });

    System.out.println("Views for articles contained by each country");
    for (LocalPage page : WpCollectionUtils.sortMapKeys(views, true)) {
        System.out.format("%s\ts\n", page.getTitle().getCanonicalTitle(), views.get(page).toString);
    }
}

private static LocalPage findCountry(Map<LocalPage, Geometry> countryShapes, Geometry point) {
    for (LocalPage country : countryShapes.keySet()) {
        if (countryShapes.get(country).contains(point)) {
            return country;
        }
    }
    return null;
}
Supply vs demand of categories

Supply:
# articles per category

Demand:
# views per category

public static void main(String args[]) throws ConfigurationException, DaoException {

   // Get the pageview dao
   Env env = EnvBuilder.envFromArgs(args);
   Language lang = env.getDefaultLanguage();
   final PageViewDao viewDao = env.getConfigurator().get(PageViewDao.class);
   final LocalCategoryMemberDao catDao = env.getConfigurator().get(LocalCategoryMemberDao.class);
   LocalPageDao pageDao = env.getConfigurator().get(LocalPageDao.class);

   // Download and import pageview stats if necessary.
   DateTime start = new DateTime(2014, 8, 14, 11, 0, 0);
   DateTime end = new DateTime(2014, 8, 14, 23, 0, 0);
   viewDao.ensureLoaded(start, end, env.getLanguages());

   // Build up set of top level categories
   final Set<LocalPage> topLevelCategories = new HashSet<>();
   LocalPage parent = pageDao.getByTitle(lang, Namespace.CATEGORY, TOP_LEVEL_PAREN);
   for (LocalPage page : catDao.getCategoryMembers(parent).values()) {
      if (page.getNameSpace().equals(Namespace.CATEGORY)) {
          topLevelCategories.add(page);
      }
   }

   // Map from page id -> num views
   final TIntIntMap allViews = viewDao.getAllViews(lang, start, end);
   final Map<LocalPage, Integer> articleCounts = new HashMap<LocalPage, Integer>();
   final Map<LocalPage, Integer> viewCounts = new HashMap<LocalPage, Integer>();
   final AtomicInteger numPages = new AtomicInteger();

   // Build up accumulators for each category by looping over pages in parallel
   ParallelForEach.iterate(
         pageDao.get(DaoFilter.normalPageFilter(lang)).iterator(),
         new Procedure<LocalPage>() {
            @Override
            public void call(LocalPage page) throws Exception {
               int views = allViews.get(page.getGlobalId());
               LocalPage cat = catDao.getClosestCategory(page, topLevelCategories, true);
               if (cat != null) {
                  if (articleCounts.containsKey(cat)) {
                     articleCounts.put(cat, articleCounts.get(cat) + 1);
                     viewCounts.put(cat, viewCounts.get(cat) + views);
                  } else {
                     articleCounts.put(cat, 1);
                     viewCounts.put(cat, views);
                  }
               }
            }
         });

   for (LocalPage page : viewCounts.keySet()) {
      System.out.format("%s\n\n", page.getTitle().getCanonicalTitle(), articleCounts.get(page),
      viewCounts.get(page),
      numPages.get());
   }
}
WikiBrain Tools IEG

Wikipedia-based algorithms

WikiBrain

Wikipedia Articles

Editors

Bots, tools, cyborgs

Wikipedia researchers
**mostSimilar?lang=simple&phrase=spider&n=3**

```json
{
    "success": true,
    "message": "",
    "diagnostics": {
        "cpuTime": 0.069754, "userTime": 0.065966
    },
    "results": [
        {
            "title": "Spider", "score": 0.9392013984939758, "lang": "simple", "articleId": 19903
        },
        {
            "title": "Arachnid", "score": 0.46658547513090154, "lang": "simple", "articleId": 22944
        },
        {
            "title": "Scorpion", "score": 0.4409242802398655, "lang": "simple", "articleId": 22045
        }
    ]
}
```

**categoriesForArticle?title=Jesus&lang=simple**

```json
{
    "success": true,
    "message": "",
    "diagnostics": {
        "cpuTime": 0.005009, "userTime": 0.003763
    },
    "article": {
        "title": "Jesus", "type": "title", "articleId": 219585
    },
    "distances": [
        {
            "distance": 0.33521585396335846, "title": "Category:Religion", "lang": "simple", "articleId": 200888
        },
        {
            "distance": 0.37135337094738713, "title": "Category:People", "lang": "simple", "articleId": 212045
        },
        {
            "distance": 0.7239022222538307, "title": "Category:Knowledge", "lang": "simple", "articleId": 212055
        },
        {
            "distance": 0.9894527716878347, "title": "Category:Science", "lang": "simple", "articleId": 212065
        },
        {
            "distance": 1.0924154851425356, "title": "Category:Geography", "lang": "simple", "articleId": 212075
        },
        {
            "distance": 1.095675386326904, "title": "Category:Everyday life", "lang": "simple", "articleId": 212085
        }
    ]
}
```
Wikipedia is a free-access, free-content Internet encyclopedia, supported and hosted by the non-profit Wikimedia Foundation.
WikiBrain developers:
Alan Morales Blanco, Margaret Giesel, Rebecca Gold, Becca Harper, Brent Hecht, Ben Hillman, Sam Horlbeck, Aaron Jiang, Matthew Lesicko, Toby Li, Yulun Li, Huy Mai, Ben Mathers, Sam Naden, Jesse Russell, Shilad Sen, Laura Sousa Vonessen, Zixiao Wang, and Ari Weilland
WikiBrain Case Studies
Cultural Alignment in Algorithms

Knowledge Base
(Wikipedia Editors)

Algorithm

Gold Standard
(Human-Labeled Data)

Application Audience
(End-Users)

Cultural Alignment in Algorithms

Localness of Sources Cited in Spatial Articles

Sen, Ford, Musicant, Graham, Keyes, Hecht. Barriers to the localness of volunteered geographic information. CHI 2015

http://shilad.com/localness
Towards domain-specific semantic relatedness: A case study from geography. Sen; Johnson; Harper; Mai; Olsen; Mathers; Vonessen; Wright; Hecht. IJCAI, 2015.

\[
GESR(A,B) = -0.019 \cdot \text{states}(A,B) + 0.173 \cdot \text{ordinal}(A,B) + 2.598 \cdot \text{general-SR}(A,B)
\]

\[
SR(\text{Minneapolis}) = 0.6
\]

\[
SR(\text{Sahara}) = 0.2
\]

Image Copyrights:
- CC0: https://www.pexels.com/photo/golden-gate-bridge-san-francisco-bay-city-29915/
- Public Domain: https://commons.wikimedia.org/wiki/File:Minneapolis_on_Mississippi_River.jpg
Thematic Cartography
Europe's Biggest Wine Drinkers

Annual per capita wine consumption in European countries (Nov 2015)*

- 45l-50l
- 30l-45l
- 20l-30l
- 10l-20l
- 0-10l

* litres per capita

Source: The Wine Institute

Tobler’s First Law of Geography: Everything is related to everything else, but near things are more related than distant things.
Spatialization
Prior Spatialization Systems

GMap
Hu et al., 2010

Atlasify
Hecht et al., 2012

Gronemann
and Jünger, 2013

...and many others!
Spatialization in Cartograph

<table>
<thead>
<tr>
<th>Title</th>
<th>Gender Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing in Action 2: The Begir</td>
<td>0.08</td>
</tr>
<tr>
<td>Cross of Iron (1977)</td>
<td>0.08</td>
</tr>
<tr>
<td>Heartbreak Ridge (1986)</td>
<td>0.08</td>
</tr>
<tr>
<td>Bloodsport (1988)</td>
<td>0.09</td>
</tr>
<tr>
<td>Predator 2 (1990)</td>
<td>0.09</td>
</tr>
<tr>
<td>Tora! Tora! Tora! (1970)</td>
<td>0.09</td>
</tr>
<tr>
<td>Red Heat (1988)</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Wikipedia
Map of Film Interest by Gender

This map visualizes the interest in films by gender. Each city represents a film, with neighboring cities being related in meaning. Film interest is defined by the relative frequency with which a gender rates a film on MovieLens.
http://cartograph.info
Cartograph Innovations:

1. Taps vast world knowledge encoded in Wikipedia.
2. Leverages recent advances in NLP algorithms.

Thematic cartography that is scalable, interactive, applicable to almost any dataset.
Cartograph Pipeline

1. Concept definition
2. X, Y embedding
3. Thematic layers
4. Map delivery
Step 1: Concept definition

Text in dataset row

NLP Algorithms
- Wikification
- Entity Recognition

Concept
- Wikidata Entity
- Popularity Estimate

Page Rank
Page Views

"Star Wars"

article in 69 langs

Importance: #5 of 72K
Berkshire Hathaway Inc. is an American multinational conglomerate holding company headquartered in Omaha, Nebraska...

Step 2: X, Y embedding

Word2Vec

Berkshire Hathaway
American
Multinational...
Holding company
Omaha, nebraska

200 columns

10 M rows

t-SNE on 50K sample points

interpolation for out of sample points
Step 2a: A Word2Vec Sentence is...

Wikipedia article text

User session article clicks
Step 3: Thematic Layers

Thematic Cluster
(Categorical Layer)

Relative Gender Interest
(Proportional Layer)
Step 4: Web Delivery

Custom Tile Server

Raster Background Tiles

Vector Foreground Data Tiles

Browser (WebGL)
Case Studies
Case Study: Map of Wikipedia

1.4 million Wikipedia articles with sufficient page views
Map of Wikipedia Topics

This map visualizes Wikipedia articles. Cities represent articles, with neighboring cities being related in meaning. Colored countries correspond to groups of related articles.
Case Study: Corporate Sustainability

Data from https://www.csrhub.com/
Exploratory Study: Feedback from Users
Goals

Learn how domain experts interpret the cartographic embedding

Learn how to design better spatialization tools to support exploratory analytical tasks
Dataset

Gender focus of articles

Ratio of links to men articles / women articles

Gender of article came from WikiData
Women's rights: Women's rights are the rights and entitlements claimed for women and girls worldwide, and formed the basis for the women's rights movement in the nineteenth century and feminist movement during the 20th century. In some countries, these rights are institutionalized or supported by law, local custom, and behavior, whereas in others they are ignored and suppressed. They differ from broader notions of...[see Wikipedia article] Article links to 57 men and 49 women.
Article Quality Visualization

Quality estimates from ORES (Halfaker)

High Quality    Low Quality

https://figshare.com/articles/Monthly_Wikipedia_article_quality_predictions/3859800
User Study Results

Overall positive feedback

Easy, fun

Some confusion about article placement
Next Steps

1. Stand up WikiBrain API in labs
2. Regular releases of navigation embeddings
3. Release Cartograph
4. Cartograph enhancements
Thank You!

Research collaborators, Wikipedians, Aaron Halfaker

http://shilad.com
http://wikibrainapi.org
http://cartograph.info